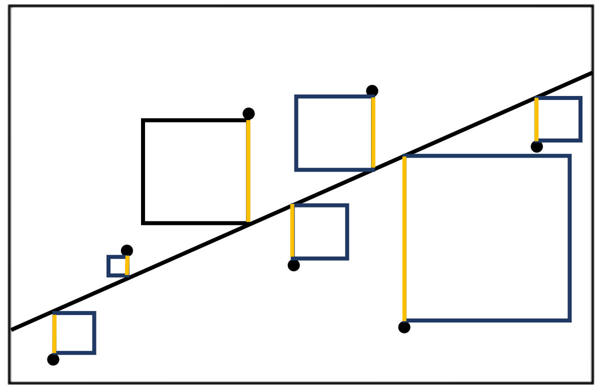
Least Absolute Deviations:

Given a data set this module aims to find the straight line that corresponds to the least absolute errors (the yellow offsets) and then compares that to its least squares variant that aims to minimize the sum of the squares instead. The least absolute deviations (LAD) approach is more robust because it gives equal emphasis to all observations, the same is not true for least squares because by minimizing the squares of the residuals it tends to give more weight to points that lie farther away from the line which translates into its sensitivity to outliers and anomalous points. However, unlike the least squares approach LAD does not have a closed for solution since the absolute function does not have a 'nice' derivative, moreover, the least squares approach seems to eclipse LAD in terms of stability, reason being that the cost function for LAD: $J = \sum|y - (a + bx_i)|$ can have more than one minimizer, all yielding the same regression error (i.e. you can often perturb the slope and the intercept of the best fit LAD line while keeping the regression error in place) the same isn't true for least squares; fitting a hyperplane will always result in a unique solution provided that the number of observations is greater than or equal to the parameters in the hyperplane.



[Fig. 1, 1] Fitting a Straight Line

Our approach to solve the LAD problem is conveyed in solving an iteratively weighted least squares problem with weight equal to $|y - (a + bx)|^{-1}$ since $\big(y - (a + bx)\big)^2|y - (a + bx)|^{-1} = |y - (a + bx)|$. Using matrix calculus one can derive the iterative scheme as follows:

$$J = (Y - X\theta)W(Y - X\theta)^t$$

Where J is the cost function, W is the weight diagonal matrix, X being the design matrix (with the observations as its rows) & Y being the solution vector.

$$(Y - X\theta)^t WX = 0 \rightarrow \theta = (X^t WX)^{-1}(X^t WY)$$

Set the derivative of J with the respect to θ to zero then solve for θ.

Inputs:

- The data set (x, y), the no. of iterations and the stopping criteria (the stopping criteria is not given as a percentage)

Outputs:

- Best fit LAD and OLS lines with their Regression Error, True Error, Correlation Coefficient respectively, the angle between the two lines.

Notes:

- The suitable norm was used in order to compute the errors of each approach (i.e. the L1 norm was used to compute the errors for LAD, the L2 norm was used for least squares.)
- Check the robustness property of LAD by adding an anomaly to your data set.
- Check the instability property of LAD by calculating again to see other solutions, notice how the regression error does not change.